Super-intelligence :
Too Smart to be Dumb or too Dumb to be Smart

Reinard Lazuardi Kuwandy
Philosophy and Artificial Intelligence 2019
Università della Svizzera italiana

April 2019

1 Introduction

(Bill Gates, 2019) *'The world hasn't had that many technologies that are both promising and dangerous — you know, we had nuclear energy and nuclear weapons.'*, *'With AI, you know the power of it is so incredible, it will change the society in very deep way'*. [1]

Since old time,Artificial Intelligence or AI, was designed to make machine smarter. Even, the idea behind of AI itself, which was also defined by The 1950s AI researchers, is by making the machine to have human-level intelligence.[2]

AI, has been developed by a lot of AI researchers in the last 40 years and now it gives a big impact on human's life. It becomes promising and well-developed on the process.

In 1997, IBM **Deep Blue**, the smart AI which build to play chess, beat Garry Kasparov $3\frac{1}{2} - 2\frac{1}{2}$ in a rematch after lost 2 - 4 before, Google Deep Mind **Alpha Go** won 4 - 1 from Go champion Lee Sedol, and recently **OpenAI** won 7,215 games with win rate 99.4% againts human in three days. Then these mean machines are smarter than human ?

In fact, above examples show that AI can still be beaten by human and there are a lot things which human can do but AI couldn't perform which means we still have control over it. Then it comes up questions, what will happen in future? When a machine keep being trained, will it surpass human intelligence? Will we always have a control over it? when we do not, what will happen ? And another random question of curiosity arises, can machine have an feeling or common sense like us ?

The answer is we do not know how machines with intelligence will turn out to be. Imagine the movie *2001: A Space Odyssey*, when HAL 9000 the super-intelligence computer,was said to be foolproof and incapable of error, suddenly turned 180 degrees, changed his behaviors by taking control of the pod and fought the astronauts who believe that the system is an error and tried to shut him down. It is so unpredictable and that is the reason why famous people like Bill Gates also feared it. It is hard to predict the behavior because it is kept

trained and learned. Probably human will not be able to regain power when it turns into super-intelligence machines, smarter than human in all domains of interest.

When it happens, fear about super-intelligent machines might be too dumb to posses common sense comes up. Indeed, if asked to "make all people happy", they might very well decide to kill all people, since trivially this would satisfies the request. Super-intelligent machine can be either too smart to be dumb or too dumb to be smart.

2 Super-intelligence

(David Chalmers, 2019)*"One of our questions here is, is superintelligence possible or impossible? I'm on the side of possible. I like the possible, which is one reason I like John's theme, "Possible Minds." That is a wonderful theme for thinking about intelligence, both natural and artificial, and consciousness, both natural and artificial. ... The space of possible minds is absolutely vast—all the minds there ever have been, will be, or could be. Starting with the actual minds, I guess there have been a hundred billion or so humans with minds of their own. Some pretty amazing minds have been in there. Confucius, Isaac Newton, Jane Austen, Pablo Picasso, Martin Luther King, on it goes. But still, those hundred billion minds put together are just the tiniest corner of this space of possible minds.[3]"*

(Nick Bostrom, 2014) Super-intelligence defines as any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest. Super-intelligence is not only AI, there are several conceivable technology paths to reach super-intelligence. We look at whole brain emulation, biological cognition, human-machine interfaces and including AI are part of technology paths to reach super-intelligence.

On this definition, **Deep Blue**, **Alpha Go** and **OpenAI** are not superintelligence since they only smart within one narrow domain and as it has been mentioned before it still can be beaten by humans and there are many fields where they perform much worse than a single human.[4]

Super-intelligence is radically different. it will lead to a more advance superintelligent. Technology that we can currently foresee will be speedily developed by the first super-intelligent. The emergence of it may happen swiftly and the possibility of it is referred to as the *singularity hypothesis.*[5]

The idea behind singularity or is known as intelligence explosion is when machine become more genious than humans, and it will be better at designing machines too. It will be capable to design machines which is more intelligent than the most perfect machines created by humans and it will turn into domino effects which means the new devices will be able to design a machine which more smart than itself.

The basic argument of singularity was set out by the statistician I. J. Good in his 1965 article "Speculations Concerning the First Ultraintelligent Machine":

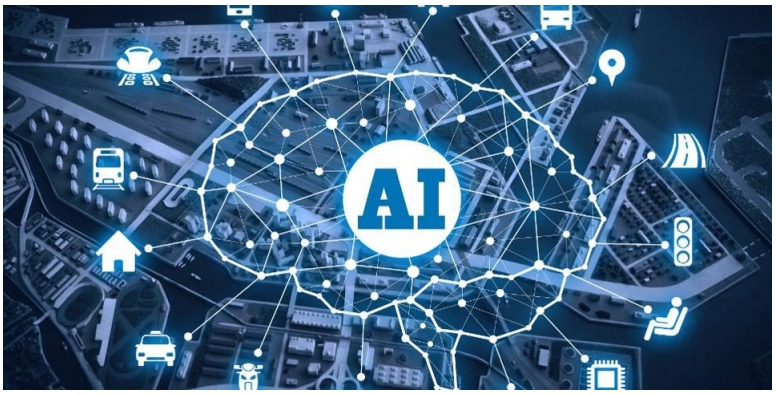

Figure 1: [illustration] AI is not Super-intelligent but it is one of the path to reach Super-Intelligent

https://cdn-images-1.medium.com/max/1200/1*zIkubEJ69fnD1CUnmDH_8g.jpeg

> Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since designing machines is one of these intellectual activities, an ultra-intelligent machine could even better machines; there would then unquestionably be an "intelligence explosion", and the intelligence of men would be left far behind. Thus the first ultra-intelligent machine is the last invention that men need ever make.[6]

This argument leads us to many philosophical questions like: how about the nature intelligence ? how about the mental capacities of artificial machines ? The potential consequences of an intelligence explosion or singularity which force us to think hard about the values, morality, consciousness and personal identification. We need to determine whether we can play a significant role in a post-singularity world, we need to better understand what intelligence is and whether machines might have it or not.

3 Super-intelligence Dilemma: Threat more than Benefit

There is no doubt, Super-intelligence could either solve or at least help us solving problems. With their intelligence and capability, problems like diseases, poverty, environmental destructions, unnecessary suffering of all kinds would be eliminated. Creating opportunities for us to enjoy our lives in which we could joyfully playing-game, relating to each other, experiencing, personal growth, and to living closer to our ideals are several other examples of benefits which are given by Super-intelligence.

On the other hand, our concerns as humans to the inevitable technology development are that we do not have control over it and nobody can guarantee if it can control itself. The fact that super-intelligence is created by several creators to favors only certain group of humans, not humanity in general. Another source of concern is the programmer teams will make a big mistake designing the goal system.

these all lead us into a state of affairs we might now like, but which in fact turns out to be a false utopia, which is essential to human flourishing have been irreversibly lost. [5] We need to be careful about what we wish for from a super-intelligence, because we might get it even more than we expect. Like when we ask super-intelligence machine to make us happy, maybe it will kill us because they don't have a common sense, they believe with what they learned and what they learned are not 100% a good things.

(Bostrom, 2002) Super-intelligence is one of several "existential risks": a risk "where an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential". Conversely, a positive outcome for superintelligence could preserve Earth-originating intelligent life and help fulfill its potential. It is important to emphasize that smarter minds pose great potential benefits as well as risks.

Kurzweil (2005) holds that "[i]ntelligence is inherently impossible to control". Let suppose that the AI is not only clever, but that, as part of the process of improving its own intelligence, it has unhindered access to its own source code: it can rewrite itself to anything it wants itself to be. Yet it does not follow that the AI must want to rewrite itself to a hostile form.

Consider Gandhi, who seems to have possessed a sincere desire not to kill people. Gandhi would not knowingly take a pill that caused him to want to kill people, because Gandhi knows that if he wants to kill people, he will probably kill people, and the current version of Gandhi does not want to kill. More generally, it seems likely that most self-modifying minds will naturally have stable utility functions, which implies that an initial choice of mind design can have lasting effects (Omohundro, 2008).[7]

This is the challenge of machine ethics: How do you build an AI in which, when it executes, becomes more ethical than you? How to create machines with consciousness, morality, and personal identify?

> Consciousness: The having of perceptions, thoughts, and feelings;an awareness. The term is impossible to define except in terms that are unintelligible, without a grasp of what consciousness means. Many fall into the trap of confusing consciousness with self-consciousness—to be conscious it is only necessary to be aware of the external world. Consciousness is a fascinating but elusive phenomenon: it is impossible to specify what it is, what it does, or why it evolves. Nothing worth reading has been written about it. (Sutherland, 1989.)[8]

It becomes a dilemma. On one side, super-intelligence gives a lot of promising benefits to help and develop human life and to upgrade human to the next

level of humanity but in the other hand it can be a weapon with their intelligence without consciousness, morality and personal identification to interpret something simple into complicated things.

4 Conclusion

Although current super-intelligence hasn't existed yet but the final stage may happen swiftly and suddenly. Humans need to be ready with it. Building a good foundation of super-intelligence with consciousness, morality and self identify is necessary.Super-intelligence will bring us to the extraordinary challenge of stating an algorithm that outputs super-ethical behavior. Context of safety assurance must be put forward. The purpose of super-intelligence should be defined, not to serve several selected group of humans, but humanity in general. In the end, we can have control of super-intelligence and indeed, if we asked to super-intelligence machine "make all people happy",they may very well decide to give all people an ice cream rather than killed us. super-intelligence machine becomes smart and hart to be dumb.

References

[1] Catherine Clifford. Bill gates: A.i. is like nuclear energy — 'both promising and dangerous', March 2019.

[2] John McCarthy. The philosophy of ai and the ai of philosophy. http://jmc.stanford.edu/articles/aiphil2/aiphil2.pdf, 2006. [Online; accessed 25-April-2019].

[3] David Chalmers, Daniel C. Dennett. Is superintelligence impossible? (on possible minds: Philosophy and ai). https://www.edge.org/conversation/david$_c halmers - daniel_{cd} ennett - is - superintelligence - impossible$, 2019. [$Online; accessed 30 - April - 2019$].

[4] Nick Bostrom. Superintelligence: Paths, dangers, strategies. http://www.korin-consulting.com/pdf/NickBostrom$_S uperintelligence.pdf$, 2014.[$Online; accessed 25 - April - 2019$].

[5] Nick Bostrom. Ethical issues in advanced artificial intelligence. https://nickbostrom.com/ethics/ai.html$_f tnref 4$. [$Online; accessed 25 - April - 2019$].

[6] David J. Chalmers. The singularity: A philosophical analysis. https://www.icorsi.ch/mod/url/view.php?id=406276. [Online; accessed 25-April-2019].

[7] Nick Bostrom, Eliezer Yudkowsky. The ethics of artificial intelligence. https://www.icorsi.ch/mod/url/view.php?id=406275, 2011. [Online; accessed 25-April-2019].

[8] David J. Chalmers. The conscious mind: In search of a theory of conscious experience. http://ghiraldelli.pro.br/wp-content/uploads/The-Conscious-Mind-Chalmers-David.pdf, 1995. [Online; accessed 30-April-2019].